A Security Scheme for Statistical Anomaly Detection and the Mitigation of Rank Attacks in RPL Networks (IoT Environment)

Mohammed A. Alqarni
Department of Software Engineering, College of Computer Science and Engineering, University of Jeddah, Saudi Arabia
alqarni@uj.edu.sa

Sajjad Hussain Chauhdary
Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering, University of Jeddah, Saudi Arabia
shussain1@uj.edu.sa (corresponding author)

ABSTRACT

A Routing Protocol for Low-power-lossy (RPL) networks builds a Destination Oriented Directed Acyclic Graph (DODAG) to provide IPv6 connectivity for resource-constrained devices over a large variety of low-power-lossy link layer technologies. Each RPL node maintains a rank value, which quantizes its relative topological distance from the DODAG root and is calculated based on the rank of its preferred parents and the objective function being employed. The RPL routing process does not impose any check to monitor the action and conduct of the parent nodes. A malicious attacking node can exploit this weakness by faking its rank value to be much lower than the original to attract more traffic to traverse through it from its neighboring and underlying child nodes. An attacking node can choose to perform selective forwarding or a sinkhole attack (Rank Attack type 1 – RA1) or exacerbate network performance parameters by causing topological instability (Rank Attack type 2 - RA2). This paper presents the Statistically-based Anomaly Detection Scheme (SARPL) to detect RA1 and RA2 and attempts to mitigate their effects. The simulations and performance evaluations show that SARPL can successfully detect RA1 attacks in all scenarios whereas it has a positive detection rate of approximately 93% for RA2 type attacks. SARPL also significantly improves network performance parameters, such as packet delivery rate and end-to-end delay, while mitigating the effects of RA1 and RA2.

Keywords - anomaly detection; rank attack; RPL network; low power lossy network

I. INTRODUCTION

Low power and Lossy Networks (LLNs) usually consist of resource constrained nodes connected using a large variety of link layer technologies, such as IoT Environment, IEEE 802.15.4, IEEE P1901.2 (Power Line Communication - PLC), IEEE 802.15.4g (Low-Power Wi-Fi) etc. A few application areas for LLNs include asset tracking, industrial and environmental monitoring, smart homes, smart energy metering, building automation and many more [1]. The lossy nature of LLNs demands a robust routing protocol that can efficiently address ephemeral and unpredictable network characteristics. IETFs ROLL (Routing Over Low-power and Lossy networks) [2, 3] working group proposed the specification of RPL (Routing Protocol over LLNs), which supports a large variety of constrained and potentially lossy link layers, suitable for resource-constrained devices. RPL constructs a logical network topology by constructing a Directed Acyclic Graph (DAG) for the communication of the network devices. RPL then divides this topology into multiple Destination Oriented Directed Acyclic Graphs (DODAGs), one for each root node, also called a sink, which is responsible for data collection and DODAG network coordination. Each DODAG RPL can be uniquely identified by a quadruple of "RPL Instance ID," "DODAG ID," "DODAG Version Number," and "Rank" as shown in Figure 1 and described in [4]. The RPL specification primarily uses four types of control messages for setting up and updating the topology of a DODAG [3]: the DIO (DODAG Information Object), DAO (DODAG Destination Object), DIS (DODAG Information Solicitation), and DAO-ACK messages.

A. Concept of Rank

Each RPL node maintains a rank value, which is calculated based on its distance from the DODAG root and indicates the quality of the path towards it. RPL nodes calculate and update
their rank value based on the rank of their preferred parents and the objective function being employed. In RPL networks, rank values strictly increase from the DODAG root towards the RPL leaf nodes to ensure an optimal network topology, prevention of routing loops and to minimize routing inconsistencies. DIO messages are used to propagate the updated rank value of a node to its neighboring nodes. The routing operation of RPL assumes that all participating nodes in the network are reliable and will follow the protocol rules strictly. A detailed discussion on rank properties, relationships and its computation can be found in [3, 5].

To launch a rank attack, a malicious node fakes its rank value based on the rank of their preferred parents and the objective function being employed. In RPL networks, rank values strictly increase from the DODAG root towards the RPL leaf nodes to ensure an optimal network topology, prevention of routing loops and to minimize routing inconsistencies. DIO messages are used to propagate the updated rank value of a node to its neighboring nodes. The routing operation of RPL assumes that all participating nodes in the network are reliable and will follow the protocol rules strictly. A detailed discussion on rank properties, relationships and its computation can be found in [3, 5].

### B. Rank Attack and its Impact

The RPL routing process does not impose any check to monitor the action of the parent nodes. The child nodes have no way to check the validity of the information provided by their parents and must depend on them exclusively. In the case of an attack, they have to rely on information provided by DIO messages disseminated by malicious nodes, resulting in non-optimized routes towards the DODAG root node. Malicious attacking nodes can exploit this weakness to launch similar attacks, such as sinkhole attacks, wormhole attacks, selective forwarding, etc. IETF RFC 7416 [6] describes some means to protect the RPL control plane against external threats to ensure its integrity, authenticity, and confidentiality (optional). However, the control plane still remains prone to intra-network attacks, as an attacker can get access to security credentials, enabling it to manipulate the RPL routing topology. The wireless sensor devices, which are usually scattered and unattended, are not tamper-resistant and are weakly secured due to their limited functionality and capabilities. One of the primary security challenges in RPL is to secure information, such as a nodes rank, which can be used for unwanted manipulation of the routing topology [7].

A study has been carried out in [4] to understand the impact of a rank attack on the topology of an RPL network. This paper outlines how a rank attack can have a severely negative and decaying performance effect on the network in terms of various parameters, such as packet delivery ratio, end-to-end delay, control traffic overhead, and network throughput. An intrusion detection system has been proposed in [10] to detect topology attacks such as rank attack. The proposed scheme requires RPL

In a rank attack, the attacker can choose to permanently violate the rank rule or to flip between being for and against it over a specific period of time [4]. Flipping is primarily used to disrupt normal network operation and to bring instability to the network topology by continually changing its preferred parent node. Thus, every time the topology around the attacking node changes due to flipping, it will generate a DIO message to update its neighbors, causing additional control overhead leading to higher energy constraints. The neighboring nodes will then be forced to update their routing information and disseminate this change by generating even more DIOs, resulting in a waste of valuable resources.

This study provides the statistically-based anomaly detection scheme SARPL that aims to detect and counteract two distinct rank attacks, RA1 and RA2. According to simulations and performance tests, SARPL has a positive detection rate of roughly 93% for RA2 type assaults and can successfully identify RA1 attacks in all scenarios. While reducing the effects of RA1 and RA2, SARPL also dramatically enhances network performance metrics like packet delivery rate and end-to-end latency.

### II. CONTRIBUTION

The proposed system is able:

- To detect and neutralize two forms of rank attacks, whereas a novel statistical anomaly detection scheme has been presented for resource-constrained RPL-based networks.
- To guarantee that a malevolent parent node does not alter or remove Rank Attack Detection (RAD) control data intended solely for the DODAG root node. To do so, a straightforward random selection of "preferred parents" system has been designed.
- To examine the effects of two distinct rank attacks, RA1 and RA2, on various network performance metrics.

Simulations of the attacks were conducted in the CONTIKI-COOJA network simulator IDE.

### III. RELATED WORK

A study has been carried out in [4] to understand the impact of a rank attack on the topology of an RPL network. This paper outlines how a rank attack can have a severely negative and decaying performance effect on the network in terms of various parameters, such as packet delivery ratio, end-to-end delay, control traffic overhead, and network throughput. An intrusion detection system has been proposed in [10] to detect topology attacks such as rank attack. The proposed scheme requires RPL
to implement a Finite State Machine (FSM) inside each node to monitor suspicious behavior. However, no simulation results were presented to validate proof of the concept. SVELTE [11] consists of a vital component named 6mapper that is used to gather topology information of the RPL network in the DODAG root node. 6mapper sends mapping requests to each RPL node in the network at regular intervals. RPL nodes respond to these requests by appending their rank, their preferred parents ID, and their neighbors IDs and respective ranks. However, SVELTE has a high false rate, which is primarily caused by valid inconsistencies in 6mapper between rank measurements [12].

The Intrusion Detection Scheme (IDS) proposed in [13] uses a simple heartbeat protocol to obtain an updated network state and subsequently to detect attacks such as sinkhole and selective forwarding. While simulating these attacks for RPL networks, it is assumed that the malicious node cannot differentiate RPL control traffic from the normal traffic and is unable to drop RPL packets. This assumption requires encryption, upper-layer security protocols such as IPSec or implementing other security mechanisms, which subsequently affects the energy efficiency of resource-constrained RPL nodes. However, if the malicious node gains access to the encryption key of the besieged nodes, this assumption is no longer workable. Authors in [14] propose a version number and VeRA rank authentication scheme to prevent such internal attacks. VeRA attempts to ensure a strict rank increase from the DODAG root towards leaf nodes by using a one-way-hash chain. However, hash chains and other such schemes are computationally expensive and not feasible for large RPL networks consisting of resource-constrained nodes [12] leading to network scalability issues. Additionally, VeRA remains vulnerable to rank attacks by forgery and replay [7]. In [15], an IDS framework (Demo) for the Internet of Things (IoT) has been proposed, where an IDS probe node is used to monitor 6LoWPAN by sniffing network traffic. Subsequently, this information is transmitted to IDS using a dedicated wired connection. The scheme requires a dedicated probe node (equipped with high computational resources and bandwidth) and a wired link to IDS for its functioning. Similarly, the network-based IDS proposed in [16] is based on multiple IDS agents spread all over the network and able to operate in promiscuous mode. Such dedicated probes act external to 6LoWPAN, do not participate in network operations and require a dedicated wired connection to the IDS system. These requirements of dedicated computational and bandwidth resources make them an unsuitable candidate for large-scale LLNs. A generic RPL topology authentication scheme that detects and prevents topological inconsistencies has been proposed by TRAIL [17]. It uses the DODAG root node as a trust anchor and monotonically increases node ranks. This scheme requires each network node to perform path validation independently, which significantly increases control messaging overhead and signature processing, further exacerbating scalability issues. TRAIL also requires each node to maintain state information, which is likely to diminish the already constrained computing and energy resources.

In [18], the goals, working methods, and advantages of the schemes are examined. There is a taxonomy based on the detection methods and a comparison table of the examined schemes. A performance study of a few classification algorithms used for network anomaly mitigation techniques in the IoT was carried out using the UNSW-NB15 dataset. The difficulties and unresolved problems in the creation of IoT network anomaly mitigation strategies were examined. Authors in [19] covered the operation of RPLs as well as a number of RPL assaults limited to LLN security. To enhance the RPL performance, an evaluation of different countermeasures against the attacks was also carried out. Authors in [20-22] considered machine learning methods for malevolent behavior detection in IoT networks.

IV. SARPL SYSTEM MODEL

In rank attacks, after comprising security credentials, a malicious node generally fakes its Rank Value to be much lower than its original value, in an attempt to attract more traffic from its neighboring nodes having higher rank values. Later on, the attacker can choose to undertake various illegal actions, such as performing selective forwarding, launch a sinkhole attack, or simply exacerbating network performance parameters. For the study of the proposed scheme, two different versions of rank attacks were simulated:

- Rank Attack 1 (RA1): The attacker intentionally drops data packets to perform selective forwarding and is also capable of randomly flipping between normal RPL operation mode and selective forwarding mode.
- Rank Attack 2 (RA2): The malicious node does not drop packets to disrupt normal network operation. Rather, it tries to affect the network performance parameters, e.g. by choosing the worst parent as its preferred parent instead of the best one [4]. Such attacks are sophisticated and difficult to detect and mitigate because, in such cases, the actions of the attackers can have significant negative impact on the network performance but very few noticeable rule violations of network procedures.

To formulate the proposed scheme, let's consider an RPL network consisting of a single DODAG G having N number of nodes. For any RPL network node ni the candidate set Cni reachable via link-local multicast, the parent set Prntni, and preferred parent (i.e. the next hop in upward routes) Prefni are represented by:

\[ C_{ni} = \{ n_j | j < N, n_j \in G, n_j \neq n_i \} \]  \hspace{1cm} (1)

\[ Prnt_{ni} = \{ n_j | P_{ni} \subseteq C_{ni}, n_j \in Same_{DODAG,Version} \} \]  \hspace{1cm} (2)

\[ Pref_{ni} = \{ n_j | n_j \in P_{ni} \} \]  \hspace{1cm} (3)

An RPL network consists mainly of resource-constrained nodes having limited computational, energy and memory resources. Alternately, the DODAG root node needs to perform duties such as acting as the anchor node (gateway) providing IPv6 connectivity with the Internet, data collection and dissemination, data processing, control plane monitoring, etc. Therefore, the root node is usually dedicatedly powered and has relatively high processing resources.

For the proposed scheme, the RPL control panel has been extended to include a control packet referred to as RAD, which
is sent periodically from nodes towards the DODAG Root Node (DRN). This enables the DRN \( D_{RN} \) to detect any anomalous behavior, such as a rank attack by malicious nodes. RAD packets have a constant packet size of 10 bytes. Their format is presented in Figure 2.

Another major issue to address is preventing malicious nodes from filtering and dropping RAD control packets of other RPL nodes traversing through it. As a deterrent against such actions, each node \( n_i \) forwards its RAD packet to a randomly selected node from its set of parent set \( Prnt_{ni} \), instead of always choosing the preferred parent node \( Pref_{ni} \). This mechanism incurs randomness to ensure resilience against malicious attacks aimed to disrupt the RPL control plane.

**A. Detection of Rank Attack 1 (RA1)**

Using a RAD packet, each RPL node \( n_i \) informs the DODAG Root Node (DRN) about the number of packets \( (p_{ni}^{sent}) \) forwarded by it to its preferred parent \( Pref_{ni} \) over the time period \( t \). This information helps \( D_{RN} \) to construct a picture of whether any particular node in the network is acting maliciously by dropping packets, forwarded by its child nodes and belonging to its sub-DODAG. For a normal (non-malicious) node, in a given time interval, the difference between packets received by it from its child nodes and packets forwarded should be zero. Otherwise, the anomaly detection scheme will mark it initially as a suspicious node.

Considering Figure 3, if an attacking node \( (n_{a}) \) tries to report an incorrect value of \( p_{ni}^{sent} \) in its RAD packet, the root node \( D_{RN} \) can calculate and evaluate in order to find this discrepancy by adding the values of the number of packets sent by its sub-DODAG to the malicious node, \( p_{n1}^{sent} + p_{n2}^{sent} + p_{n3}^{sent} + p_{n7}^{sent} \), and the number of packets forwarded by the attacker, \( p_{n6}^{sent} \) as presented in (4):

\[
\theta_{ni}(t) = \sum_{child=1}^{N-1} p_{child}^{sent} - p_{ni}^{sent} \tag{4}
\]

where \( \theta_{ni}(t) \) is the difference in the number of packets over the time period \( t \). For any node \( n_i \), if \( \theta_{ni}(t) = 0 \), the node is implied to be operating normally.

**B. Detection of Rank Attack 2 (RA2)**

In the second type of rank attack, RA2, the malicious node does not drop packets. Rather, it negatively impacts network performance by forwarding data packets of the child nodes toward the worst-performing parent instead of the best one. As a result, an additional delay will incur in the end-to-end transmission time of packets. When sending RAD packets towards the root node, each RPL node \( n_i \) randomly selects the next hop from the available parent set \( Prnt_{ni} \).

Consider a node \( n_i \) whose parent is a malicious node carrying out the second version of a rank attack. In this case, some of the nodes RAD packets will reach the root node by transiting through the malicious node, while others will arrive through normally operating nodes. The root node can find the difference between the travel times of incoming RAD packets to detect anomalous behavior of any nodes parent, as represented in (5)-(8).

\[
\omega_{ni} = |Send_{time} - Arrival_{time}| \tag{5}
\]

where \( \omega_{ni} \) = end-to-end delay of node's \( n_i \)'s RAD packet, \( Send_{time} \) = value of the RAD packets time stamp, \( Arrival_{time} \) = arrival time of the RAD packet at the DODAGRoot node

\[
S_t = \sum_{k=1}^{t} \omega_{ni}^k/k \tag{6}
\]

For \( t > k \),

\[
S_t = \alpha \omega_{ni}^t + (1 - \alpha)S_{t-1} \tag{7}
\]

In (6), \( S_t \) represents the value of exponential moving average at any time period \( t \). To calculate \( S_t \), the value of \( t \) is assumed to be 10, which is equal to the number of supervised RAD control packet arrivals at the start of the simulation, at which time the attack is assumed to have not yet occurred. \( S_t \) represents the exponentially moving average value of packet travel time at any time period \( t \). In (7), the coefficient \( \alpha \) characterizes the degree of weighting decrease. It is a constant smoothing factor having a value between 0 and 1. A larger value of \( \alpha \) is used to discount older observations relatively quicker. For simulations, the value of \( \alpha \) is assumed to be 0.6. For any RPL node \( n_i \), if the condition given in (8) is true, that node is marked as suspicious.

\[
\omega_{ni}^t > S_t + 0.1 * S_t \tag{8}
\]

If any node is marked as suspicious 3 consecutive times, it is considered a malicious node by the DODAG root node. In this case, all RPL nodes are notified to not use it as a preferred parent and to include it in a blacklist of restricted nodes. The pseudocode for the proposed anomaly detection and mitigation scheme for the rank attacks is presented in Figure 4.

**V. RESULTS AND DISCUSSION**

**A. Detection (True Positive) Rate**

The efficiency of any intrusion detection can be quantized in terms of its detection rates. Detection Rate (True Positive) is here defined as the number of malicious RPL nodes that have
been correctly detected by SARPL divided by the total number of attacking RPL nodes in the network. Three scenarios were considered: when all attacking nodes are of type RA1, when all attacking nodes are of type RA2, and when there is an equal number of attacking nodes of both types. A comparative analysis of SARPL in terms of detection rate is presented in Figure 5.

![Fig. 4. Pseudocode for the anomaly detection and mitigation scheme.](image)

It clearly reflects that SARPL can successfully detect RA1-type (packet dropping) attacks for all the considered numbers of attacking nodes in the network. It also shows that SARPL performs better when dealing with RA1 than with RA2. The main reason is that in a RA2 type attack, the malicious node abides by all the rules of the RPL packet routing procedures except that it chooses the worst parent instead of the best as the next hop for upstream packet forwarding. In the simulation, the attacking nodes are deployed randomly over the network topology (grid). Thus, if the attacking node is already closer to the DODAG root node, even choosing the worst parent will allow it to escape SARPL’s condition (8) for detection of RA2-type rank attacks.

**B. Packet Delivery Ratio (PDR)**

PDR is an important parameter for analyzing the impact of various security attacks on network performance. PDR is defined here as the total number of packets received by the DODAG root node divided by the total number of data packets generated by all non-root RPL nodes. Figure 6 represents the effect of a rank attack along with selective forwarding (RA1) on RPLs PDR. It also describes the performance of the SARPL scheme against security attacks that involves packet dropping by malicious nodes. The graphical representation in Figure 6 shows that a RA1-type attack, which involves a rank attack to attract traffic from its sub-DODAG and then the launch of selective forwarding to drop packets destined for DRN, can have a significant impact on RPLs’ PDR. The simulation results show that an RPL network of 50 nodes having a grid topology, with 20% of the nodes being malicious that execute RA1 attacks cause the networks PDR to drop to approximately 50%. SARPL has a 100% detection ratio for attacks such as RA1. However, when a node is marked as malicious and is blacklisted from use as a preferred parent, the rest of the network nodes are left with fewer nodes to be used as the next hop towards DRN. In the worst-case scenario, this can cause holes in the network.

![Fig. 5. SARPL detection rate evaluation.](image)

![Fig. 6. Performance comparison of the SARPL scheme against rank attacks followed by selective forwarding and packet dropping.](image)

The second type of rank attack (RA2) discussed in this paper does not involve selective forwarding or the dropping of data packets. Rather, the malicious node chooses the worst parent instead of the best one to impact network performance parameters, such as end-to-end packet delay. Figure 7 shows...
the impact of RA2 on the average end-to-end delay of the RPL network and the performance of the SARPL scheme to mitigate its effect as proof of concept. The plot represented in Figure 7 shows that, with an increasing number of attacking nodes acting maliciously, their combined effect on the average end-to-end delay increases too. SARPL improves end-to-end delay performance of the network up to an average of 32%, when 20% of the nodes start to act maliciously. However, any attacking node will go undetected, that incurs an additional end-to-end delay, lesser than detection-threshold represented by (8).

VI. CONCLUSION

The Statistical-based Anomaly Detection Scheme SARPL has been proposed and implemented to detect and mitigate the effects of rank attacks on various network performance parameters. Two different kinds of rank attack, RA1 and RA2, were simulated and examined for the purpose of this study.

The proposed scheme shown successful detection of RA1 attacking nodes and has up to 93% positive-true detection rate for RA2 type attacks. Anomaly detection of RA2 is relatively difficult compared to RA1 because the malicious nodes abide to all the rules of RPL packet routing procedures except that they choose the worst parent instead of the best one. SARPL also shows significant improvement in network performance parameters such as packet delivery ratio and end-to-end delay, while mitigating the effects of RA1 and RA2. For example: when 20% of the network nodes start to act maliciously, SARPL exhibit an average improvement of approximately 41% in packet delivery ratio and 32% in end-to-end packet delay performance.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia, for funding this research work through the project number MoE-IF-G-20-14

REFERENCES


